One-shot Slepian-Wolf

Naresh Sharma[1, *] and Naqueeb Ahmad Warsi[1, †]

[1]*School of Technology and Computer Science, Tata Institute of Fundamental Research (TIFR), Mumbai 400 005, India*

(Dated: January 5, 2012)

Slepian and Wolf gave the rate-region for the distributed encoding of correlated and memoryless sources when the number of copies of source output is unlimited. We give one-shot rate region for the Slepian-Wolf protocol when a single copy of the source output is available. Our results are asymptotically optimal, i.e., they yield the same rate region as the Slepian-Wolf in the limit of unlimited copies. We also give an upper bound for the one-shot encoding of a single source that is different from the one given by Renner and Wolf.

I. INTRODUCTION

Over the last several decades, information theory has been used to analyze the performance of various information processing tasks such as data compression, information transmission across a noisy communication channel etc. The analysis has traditionally been asymptotic wherein a certain task is repeated unlimited number of times. Such an approach has yielded rich dividends providing fundamental limits to the performance and gave operational meaning to quantities such as entropy and mutual information [1, 2].

But the assumption of repeating an information processing task is hardly realistic. There has been a great interest over the last few years to analyze a task done only *once*. Typically, the analysis involves finding lower and upper bounds on the resources needed to perform such a task. This has been referred to as “one-shot” or “single-shot” in the literature.

The lower and upper bounds are expected to be asymptotically tight, i.e., they both yield the same average number of resources needed per task when the number of times a task is performed is unbounded and this quantity is also equal to the asymptotic analysis that has been done traditionally. It is interesting to note that while the asymptotic analysis would start with the asymptotic equipartition property (AEP), the one-shot analysis could, at the very end, be augmented with AEP to yield the same answer. Furthermore, one-shot analysis has been applied to the cases where i.i.d. (independent and identically distributed) and the memoryless assumption is not valid. For example, in a data compression task, we may not have i.i.d. copies of the random variable modeling the source output or in a communication across a noisy channel, the channel may not be memoryless.

The one-shot bounds are typically in terms of smooth Rényi entropies (defined later). The elegant notion of smooth Rényi entropies were introduced by Renner and Wolf in Refs. [3, 4] where they also gave one-shot bounds for data compression and randomness extraction (see also Ref. [5]). Another important application of smooth Rényi entropy measure has been the one-shot bounds for the channel capacity given by Renner, Wolf and Wullschleger [6]. Channel coding bounds are also derived in Ref. [7] using a quantity called the smooth 0-divergence, which is a generalization of Rényi’s divergence of order 0.

*Electronic address: nsharma@tifr.res.in
†Electronic address: naqueeb@tifr.res.in

The smooth Rényi entropies were extended to the quantum case by Renner and König [8]. There has been a considerable work on the one-shot bounds for the quantum case under various scenarios (see for example Refs. [9–14] and references therein).

In Ref. [15], Schoenmaker *et al* introduce the notion of smooth Rényi entropy for the case of stationary ergodic information sources, thereby generalizing previous work which concentrated mainly on i.i.d. information sources. In Ref. [16], Holenstein and Renner give explicit and tight bounds on the smooth entropies of n-fold product distributions in terms of the Shannon entropy of a single distribution. Renes and Renner consider the problem of classical data compression in the presence of quantum side information in Ref. [17].

It is worth noting that for the classical case, to the best of authors' knowledge, there is no one-shot result in multi-party scenario, i.e., where there are multiple parties at the sender and/or receiver. Let us concentrate on the distributed encoding (data compression) protocol of correlated information sources. In a very remarkable and fundamental paper, Slepian and Wolf showed that there is no loss in the compression efficiency for the distributed encoding as compared to the collaborative encoding [18].

The fundamental contributions by Renner and Wolf [3, 4] have not been extended thus far to the Slepian-Wolf distributed data compression protocol. The upper bound that Renner and Wolf derive for their data compression comes in the form of two steps. One is the encoding (that can be accomplished by two-universal hash functions) and second is an elegant step of bootstrapping the first step by encoding a random variable that is in a ball (appropriately defined) around the random variable that we want to compress. Such an approach distributes the error in the two steps in a flexible way and yields an upper bound that is asymptotically tight. To carry over this approach to the Slepian-Wolf protocol would involve proving the existence of more than one random variables that satisfy certain conditions and the authors know of no such proof in the literature.

In this paper, we let go of the second step and define a set for the one-shot case reminiscent (though not the same) of the typical set in the asymptotic case. We don't demarcate the lines between the various steps and our encoding (based on two-universal hash functions) together with this set come together to yield a one-shot characterization of the Slepian-Wolf protocol. Interestingly, our bounds involves smooth Rényi entropy of the order $-\infty$. Rényi entropies are typically defined of order $\alpha \geq 0$ and appropriate limits have to be taken for $\alpha = 1$ that corresponds to the Shannon entropy [19]. In the definition of smooth Rényi entropy, we let $\alpha \in [-\infty, \infty]$.

II. PRIOR ONE-SHOT RESULTS ON DATA COMPRESSION

We first give the definitions of the smooth entropies that are needed for the one-shot bounds.

A. Definition of various entropies

We shall assume that all random variables in this paper are discrete and take values over a finite set. Let X be a random variable taking values over alphabet $\mathcal{X}$ with probability mass function (PMF) $P_X(x)$, $x \in \mathcal{X}$. Then the Shannon entropy [1] $H(X)$ of X is defined by

$$H(X) := -\sum_{x\in\mathcal{X}} P_X(x)\log[P_X(x)], \tag{1}$$

where we shall assume that the log is to the base 2 throughout this paper. The Rényi entropy [19] of X is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha}\log\left[\sum_{x\in\mathcal{X}} P_X^\alpha(x)\right], \tag{2}$$

where $\alpha \in [0,\infty]$ and appropriate limits are taken for $\alpha = 1$. In particular, for $\alpha = 0$, the Rényi entropy is given by

$$H_0(X) := \log|\{x \in \mathcal{X} : P_X(x) > 0\}| \tag{3}$$

and for $\alpha = \infty$, the Rényi entropy is given by

$$H_\infty(X) := -\log\left[\max_{x\in\mathcal{X}} P_X(x)\right]. \tag{4}$$

Let X, Y be two random variables with joint distribution P_{XY}, the conditional Rényi entropy of order α is defined as

$$H_\alpha(X|Y) := \frac{1}{1-\alpha}\log\left[\max_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_{X|Y=y}^\alpha(x)\right]. \tag{5}$$

The ε-smooth conditional Rényi entropy of order α [3] is defined for $\varepsilon \geq 0$ as

$$H_\alpha^\varepsilon(X|Y) := \frac{1}{1-\alpha}\log\left[\inf_{\bar{X}\bar{Y}:\Pr[\bar{X}\bar{Y}\neq XY]\leq\varepsilon}\ \max_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_{\bar{X}|\bar{Y}=y}^\alpha(x)\right]. \tag{6}$$

Although the Rényi entropy is typically defined for $\alpha \in [0,\infty]$, we let $\alpha \in [-\infty,\infty]$ and, in particular, consider the Rényi entropy of order $\alpha = -\infty$ given by

$$H_{-\infty}(X|Y) := -\log\left[\min_{y\in\mathcal{Y}}\ \min_{x:P_{X|Y=y}(x)>0} P_{X|Y=y}(x)\right] \tag{7}$$

and its smooth version for $\varepsilon \geq 0$ as

$$H_{-\infty}^\varepsilon(X|Y) := -\log\left[\sup_{\bar{X}\bar{Y}:\Pr[\bar{X}\bar{Y}\neq XY]\leq\varepsilon}\ \min_{y\in\mathcal{Y}}\ \min_{x:P_{\bar{X}|\bar{Y}=y}(x)>0} P_{\bar{X}|\bar{Y}=y}(x)\right]. \tag{8}$$

We shall use $H_\alpha^\varepsilon(X)$ interchangeably with $H_\alpha^\varepsilon(P_X)$.

B. Prior results

Let (X,Y) be two random variables taking values over alphabet $\mathcal{X}\times\mathcal{Y}$ with joint distribution P_{XY}. We shall denote this by $(X,Y) \sim P_{XY}$. X is available with Alice at location A and she wants to send it to Charlie at location C who has an error-free copy of Y. Alice encodes X and sends it over a noiseless channel to Charlie who decodes it. Our goal is to construct a protocol to minimize the number of distinct encoder

outputs while keeping the probability of the decoder output being in error (or not equal to the encoder input) to be within some predefined limits.

To accomplish this, Renner and Wolf came up with the following protocol [4]. Alice applies a random function (encoder) to X and then sends it to Charlie who has the knowledge of the random function used by the encoder and applies a function depending on the encoder and Y. If the probability of error averaged over all random functions is small, then there must exist at least one pair of functions (or the encoder/decoder pair) which shall accomplish our task. This is stated formally as follows.

Definition 1 (Renner and Wolf [4])**.** *For $(X,Y) \sim P_{XY}$ and the error ε, $0 \leq \varepsilon \leq 1$, let $\Lambda_\varepsilon^P(\mathcal{X} \to \mathcal{C})$ denote the set (e,R) where R is a random variable with range $\mathcal{R}$, $e : \mathcal{X} \times \mathcal{R} \to \mathcal{C}$ such that there exists a decoding function $g : \mathcal{C} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{X}$ with*

$$\Pr\left\{\hat{X}(R) \neq X\right\} \leq \varepsilon, \tag{9}$$

where

$$\hat{X}(R) := g\left[e(X,R), Y, R\right]. \tag{10}$$

The definition for the minimum encoding length is given by

$$\ell_{\text{enc}}^{\varepsilon}(X|Y) := \min_{\mathcal{C}:\exists(e,R)\in\Lambda_\varepsilon^P(\mathcal{X}\to\mathcal{C})} \log|\mathcal{C}|. \tag{11}$$

$$\tag{12}$$

Theorem 1. *(Renner and Wolf [4]) Let $(X,Y) \sim P_{XY}$ and ε, ε_0 and $\varepsilon_1 \in \mathbb{R}^+$ with $\varepsilon_0 + \varepsilon_1 \leq \varepsilon$, then*

$$H_0^{\varepsilon}(X|Y) \leq \ell_{\text{enc}}^{\varepsilon}(X|Y) \leq H_0^{\varepsilon_1}(X|Y) - \log(\varepsilon_2). \tag{13}$$

III. Distributed Encoding of Correlated Sources

We describe the task of distributed encoding of correlated sources in this section.

Let $(X,Y) \sim P_{XY}$. Assume that the random variable X is available with Alice at a location A and the random variable Y is available with Bob at a separate location B. Both Alice and Bob want to get across the pair (X,Y) to Charlie at location C without collaborating with each other within some prescribed error. We state this formally as follows.

Definition 2. *For the error ε, $0 \leq \varepsilon \leq 1$, let $\Lambda_\varepsilon^P(\mathcal{X} \times \mathcal{Y} \to \mathcal{C}_\mathcal{X} \times \mathcal{C}_\mathcal{Y})$ denote the set $(e_\mathcal{X}, e_\mathcal{Y}, R)$ where R is a random variable with range $\mathcal{R}$, $e_\mathcal{X} : \mathcal{X} \times \mathcal{R} \to \mathcal{C}_\mathcal{X}$ and $e_\mathcal{Y} : \mathcal{Y} \times \mathcal{R} \to \mathcal{C}_\mathcal{Y}$ such that there exists a decoding function $g : \mathcal{C}_\mathcal{X} \times \mathcal{C}_\mathcal{Y} \times \mathcal{R} \to \mathcal{X} \times \mathcal{Y}$ with*

$$\Pr\left\{(\hat{X}(R), \hat{Y}(R)) \neq (X,Y)\right\} \leq \varepsilon, \tag{14}$$

where

$$(\hat{X}(R), \hat{Y}(R)) := g\left[e_\mathcal{X}(X,R), e_\mathcal{Y}(Y,R), R\right]. \tag{15}$$

Occasionally, we shall also use the following

$$g_{\mathcal{X}}\left[e_{\mathcal{X}}(X,R),e_{\mathcal{Y}}(Y,R),R\right]:=\hat{X}(R), \tag{16}$$

$$g_{\mathcal{Y}}\left[e_{\mathcal{X}}(X,R),e_{\mathcal{Y}}(Y,R),R\right]:=\hat{Y}(R). \tag{17}$$

The definitions for the minimum encoding lengths are given by

$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X):=\min_{\mathcal{C}_{\mathcal{X}}:\exists(e_{\mathcal{X}},e_{\mathcal{Y}},R)\in\Lambda^P_{\varepsilon}}\log|\mathcal{C}_{\mathcal{X}}|, \tag{18}$$

$$\ell^{\varepsilon}_{\mathrm{d-enc}}(Y):=\min_{\mathcal{C}_{\mathcal{Y}}:\exists(e_{\mathcal{X}},e_{\mathcal{Y}},R)\in\Lambda^P_{\varepsilon}}\log|\mathcal{C}_{\mathcal{Y}}|, \tag{19}$$

where we have written Λ^P_{ε} for $\Lambda^P_{\varepsilon}(\mathcal{X}\times\mathcal{Y}\to\mathcal{C}_{\mathcal{X}}\times\mathcal{C}_{\mathcal{Y}})$.

IV. MAIN RESULTS

In this section, we summarize the results and the proofs are given in the next section.

We first need the following definition.

Definition 3. *For $(X,Y)\sim P_{XY}$, we define the following sets*

$$\mathcal{A}_{\delta}(P_X):=\{x:\Xi^{\delta}_{\min}(P_X)\le-\log P_X(x)\le\Xi^{\delta}_{\max}(P_X)\}, \tag{20}$$

where

$$\Xi^{\delta}_{\min}(P_X):=H(X)-|H(X)-H^{\delta}_{\infty}(X)|-\delta\log|\mathcal{X}|, \tag{21}$$

$$\Xi^{\delta}_{\max}(P_X):=H(X)+|H(X)-H^{\delta}_{-\infty}(X)|+\delta\log|\mathcal{X}|, \tag{22}$$

$$\mathcal{A}_{\delta}(P_Y):=\{y:\Xi^{\delta}_{\min}(P_Y)\le-\log P_Y(y)\le\Xi^{\delta}_{\max}(P_Y), \tag{23}$$

$$\begin{aligned}\mathcal{A}_{\delta}(P_{XY}):=\{(x,y):\ &\Xi^{\delta}_{\min}(P_{XY})\le-\log P_{XY}(x,y)\le\Xi^{\delta}_{\max}(P_{XY}),\\ &x\in\mathcal{A}_{\delta}(P_X),y\in\mathcal{A}_{\delta}(P_Y)\}.\end{aligned} \tag{24}$$

Our first result is to give an upper bound different from the one in Theorem 1 and we don't change the lower bound but it is mentioned for completeness.

Theorem 2. *Let $(X,Y)\sim P_{XY}$ and $\varepsilon,\varepsilon_0,\varepsilon_1\in\mathbb{R}^+$ with $\varepsilon_0+\varepsilon_1\le\varepsilon$, then*

$$H^{\varepsilon}_0(X|Y)\le\ell^{\varepsilon}_{\mathrm{enc}}(X|Y)\le\Xi^{\delta}_{\max}(P_{XY})-\Xi^{\delta}_{\min}(P_Y)-\log\varepsilon_1, \tag{25}$$

where $\delta\ge0$ is chosen such that

$$\Pr\{X\notin\mathcal{A}_{\delta}(P_X)\}\le\varepsilon_0. \tag{26}$$

We give the one-shot bounds for the distributed encoding for correlated sources in the following theorem.

Theorem 3. *Let* $(X,Y) \sim P_{XY}$ *and* $\varepsilon, \varepsilon_i \in \mathbb{R}^+$, $i = 0,1,2,3$ *with* $\sum_{i=0}^{3} \varepsilon_i \leq \varepsilon$.
Lower bounds*: For any distributed encoding protocol, the following lower bounds shall hold*

$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X) \geq H^{\varepsilon}_0(X|Y), \tag{27}$$
$$\ell^{\varepsilon}_{\mathrm{d-enc}}(Y) \geq H^{\varepsilon}_0(Y|X) \tag{28}$$
$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X) + \ell^{\varepsilon}_{\mathrm{d-enc}}(Y) \geq H^{\varepsilon}_0(X,Y). \tag{29}$$

Achievability*: There exists a distributed encoding protocol for a rate pair* $(\ell^{\varepsilon}_{\mathrm{d-enc}}(X), \ell^{\varepsilon}_{\mathrm{d-enc}}(Y))$ *satisfying the following conditions*

$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X) \geq \Xi^{\delta}_{\max}(P_{XY}) - \Xi^{\delta}_{\min}(P_Y) - \log \varepsilon_1, \tag{30}$$
$$\ell^{\varepsilon}_{\mathrm{d-enc}}(Y) \geq \Xi^{\delta}_{\max}(P_{XY}) - \Xi^{\delta}_{\min}(P_X) - \log \varepsilon_2, \tag{31}$$
$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X) + \ell^{\varepsilon}_{\mathrm{d-enc}}(Y) \geq \Xi^{\delta}_{\max}(P_{XY}) - \log \varepsilon_3, \tag{32}$$

where δ *is chosen such that*

$$\Pr\left\{(X,Y) \notin \mathcal{A}_{\delta}(P_{XY})\right\} \leq \varepsilon_0. \tag{33}$$

Next we show that the one-shot bounds given in Theorem 2 and one-shot rate region in Theorem 3 are asymptotically optimal, i.e., both the lower bounds and the achievable rates (normalized appropriately) are the same when the number of i.i.d. copies of (X,Y) pairs grows unbounded and the errors become arbitrarily small, and that this rate region in the distributed encoding case comes out to be the same as that of Slepian-Wolf [18]. Let

$$(\mathbf{X}^n_1, \mathbf{Y}^n_1) := [(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)], \tag{34}$$

where $(X_1,Y_1),...(X_n,Y_n)$ are n i.i.d. pairs of random variables distributed according to P_{XY}.

Lemma 1. *The normalized asymptotic limit of* (25) *is given by*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell^{\varepsilon}_{\mathrm{enc}}(\mathbf{X}^n_1|\mathbf{Y}^n_1)}{n} = H(X_1|Y_1). \tag{35}$$

Similarly, we can get the asymptotic limit for Theorem 3 which yields the rate-region of Slepian-Wolf [18].

Lemma 2. *The normalized asymptotic limits for the bounds in Theorem 3 are given by*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell^{\varepsilon}_{\mathrm{d-enc}}(\mathbf{X}^n_1)}{n} \geq H(X_1|Y_1), \tag{36}$$
$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{\ell^{\varepsilon}_{\mathrm{d-enc}}(\mathbf{Y}^n_1)}{n} \geq H(Y_1|X_1), \tag{37}$$
$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \left[\frac{\ell^{\varepsilon}_{\mathrm{d-enc}}(\mathbf{X}^n_1) + \ell^{\varepsilon}_{\mathrm{d-enc}}(\mathbf{Y}^n_1)}{n}\right] \geq H(X_1,Y_1). \tag{38}$$

V. PROOFS OF THE RESULTS

Some of the proofs rely upon the two-universal hash functions and we give their definition first (see also Ref. [20] and references therein).

Definition 4. *A random function $f : \mathcal{X} \times \mathcal{R} \to \mathcal{C}$ takes an input $x \in \mathcal{X}$ and generates a uniform random variable R taking values over alphabet $\mathcal{R}$ and outputs $c \in \mathcal{C}$. f is called a two-universal hash function if for any $x \neq \acute{x}$, $x, \acute{x} \in \mathcal{X}$, we have*

$$\Pr\left\{f(x,R) = f(\acute{x},R)\right\} \leq \frac{1}{|\mathcal{C}|}. \tag{39}$$

We prove the following lemma that shall be needed for the probability of error analysis later.

Lemma 3. *Let $(X,Y) \sim P_{XY}$. Define*

$$\mathcal{A}_\delta(P_{X|Y=y}) := \{x : (x,y) \in \mathcal{A}_\delta(P_{XY})\}. \tag{40}$$

For any $\delta \geq 0$, $y \in \mathcal{A}_\delta(P_Y)$, we have

$$|\mathcal{A}_\delta(P_{X|Y=y})| \leq 2^{\Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_Y)}. \tag{41}$$

Proof. For any $(x,y) \in \mathcal{A}_\delta(X,Y)$, we have

$$P_{X|Y=y}(x) = \frac{P_{XY}(x,y)}{P_Y(y)} \geq 2^{-\Xi^\delta_{\max}(P_{XY}) + \Xi^\delta_{\min}(P_Y)}. \tag{42}$$

The lemma now follows straightforwardly from

$$1 \geq \sum_{x \in \mathcal{A}_\delta(X|Y=y)} P_{X|Y=y}(x) \geq |\mathcal{A}_\delta(X|Y=y)| 2^{-\Xi^\delta_{\max}(P_{XY}) + \Xi^\delta_{\min}(P_Y)}, \tag{43}$$

where we have used (42). □

The following properties of $H^\varepsilon_{-\infty}$ are needed later in the paper.

Lemma 4. *The following holds for any $\varepsilon > 0$*

$$H^\varepsilon_{-\infty}(X) \geq H^\varepsilon_0(X). \tag{44}$$

Proof. For all $\delta > 0$, there exists a random variable $\bar{X}$ with $\Pr\{X \neq \bar{X}\} \leq \varepsilon$ such that

$$H^\varepsilon_{-\infty}(X) \geq H_{-\infty}(\bar{X}) - \delta. \tag{45}$$

Now the following inequalities hold

$$\begin{aligned} H^\varepsilon_{-\infty}(X) &\geq H_{-\infty}(\bar{X}) - \delta && (46)\\ &\geq H_0(\bar{X}) - \delta && (47)\\ &\geq H^\varepsilon_0(X) - \delta. && (48) \end{aligned}$$

Since this is true for any $\delta > 0$, the result follows. □

Lemma 5. *If $X_1^n = [X_1, ..., X_n]$ be n i.i.d. random variables distributed according to P_X, then*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{H_{-\infty}^{\varepsilon}(X_1^n)}{n} = H(X_1). \tag{49}$$

Proof. The typical set, $\mathcal{T}_\varepsilon^{(n)} \subseteq \mathcal{X}^n$, $\varepsilon \geq 0$ (see Ref. [2] for details) is defined as

$$\mathcal{T}_\varepsilon^{(n)} = \left\{ x^n : \left| -\frac{1}{n} \log P(X_1^n) - H(X_1) \right| \leq \varepsilon \right\}. \tag{50}$$

We now define a random variable Z as

$$Z = \begin{cases} X_1^n & \text{if } X_1^n \in \mathcal{T}_\varepsilon^{(n)} \\ W & \text{if } X_1^n \notin \mathcal{T}_\varepsilon^{(n)} \end{cases} \tag{51}$$

where W is uniformly chosen at random from the set $\mathcal{T}_\varepsilon^{(n)}$. We now have

$$\Pr\{Z \neq X_1^n\} = \Pr\{X_1^n \notin \mathcal{T}_\varepsilon^{(n)}\} \tag{52}$$

$$\leq \varepsilon, \tag{53}$$

where we have assumed that n is large enough so that (53) is satisfied. To get a bound on $\min_z \Pr\{Z = z\}$, we note that for $w \in \mathcal{T}_\varepsilon^{(n)}$, $\Pr\{X_1^n = w\} \geq 2^{-n(H(X_1)+\varepsilon)}$. Now using the definition of $H_{-\infty}^{\varepsilon}(X)$ and $\Pr\{Z \neq X_1^n\} \leq \varepsilon$, we have $H_{-\infty}^{\varepsilon}(X_1^n) \leq H_{-\infty}(Z)$, and hence,

$$\frac{H_{-\infty}^{\varepsilon}(X_1^n)}{n} \leq \frac{H_{-\infty}(Z)}{n} \leq H(X_1) + \varepsilon, \tag{54}$$

or

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{H_{-\infty}(X_1^n)}{n} \leq H(X_1). \tag{55}$$

We now use Lemma 4 to get $H_{-\infty}^{\varepsilon}(X_1^n) \geq H_0^{\varepsilon}(X_1^n)$ and hence,

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{H_{-\infty}^{\varepsilon}(X_1^n)}{n} \geq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{H_0^{\varepsilon}(X_1^n)}{n} = H(X_1), \tag{56}$$

where the equality in (56) follows from Ref. [4]. Thus, from (55) and (56), the lemma is proved. □

A. Proof of upper bound in Theorem 2

We show the existence of a communication protocol based on two-universal hash functions and is given by the following steps.

1. Let $e : \mathcal{X} \times \mathcal{R} \to \mathcal{C}$ be a two-universal hash function and let $\mathcal{C} = \{1, 2, \cdots, 2^\ell\}$. Alice takes $x \in \mathcal{X}$, generates R (known to Charlie), and sends $i = e_{\mathcal{X}}(x, R)$ to Charlie.

2. Charlie passes the received index i to the decoder $g : \mathcal{C} \times \mathcal{Y} \times \mathcal{R} \to \mathcal{X}$ that outputs $\hat{x} = x$ if there is only one x such that $e(x, R) = i$ and $(x, Y) \in \mathcal{A}_\delta(P_{XY})$. In all other cases, it declares an error.

Let $\varepsilon_0, \varepsilon_1 \in \mathbb{R}^+$ and $\varepsilon_0 + \varepsilon_1 \leq \varepsilon$. The probability of error of the above protocol is computed by defining the following events.

$$E_0 := \{(X, Y) \notin \mathcal{A}_\delta(P_{XY})\}, \tag{57}$$
$$E_1 := \{\exists\, \acute{x} \neq X : e(\acute{x}, R) = e(X, R), (\acute{x}, Y) \in \mathcal{A}_\delta(P_{XY})\}. \tag{58}$$

The probability of error is given by the probability of the union of these events and we upper bound that by the union bound as

$$P_e = \Pr\{E_0 \cup E_1\} \tag{59}$$
$$\leq \Pr\{E_0\} + \Pr\{E_1\}. \tag{60}$$

We could choose a $\delta \geq 0$ s.t. $\Pr\{E_0\} \leq \varepsilon_0$. Such a choice of δ clearly exists since by choosing $\delta = 0$, $\Pr\{E_0\} = 0$. We now bound $\Pr\{E_1\}$ as follows.

$$\Pr\{E_1\} = \Pr\{\exists\, \acute{x} \neq X : e(\acute{x}, R) = e(X, R), (\acute{x}, Y) \in \mathcal{A}_\delta(P_{XY})\} \tag{61}$$
$$= \sum_{x,y} P_{XY}(x, y) \Pr\{\exists\, \acute{x} \neq x : e(\acute{x}, R) = e(x, R), (\acute{x}, y) \in \mathcal{A}_\delta(P_{XY})\} \tag{62}$$
$$= \sum_{x,y} P_{XY}(x, y) \sum_{\acute{x} \neq x, (\acute{x}, y) \in \mathcal{A}_\delta(P_{XY})} \Pr\{e(\acute{x}, R) = e(x, R)\} \tag{63}$$
$$\stackrel{a}{\leq} \sum_{x,y} P_{XY}(x, y) 2^{-\ell} |\mathcal{A}_\delta(P_{X|Y=y})| \tag{64}$$
$$\stackrel{b}{\leq} 2^{\Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_Y) - \ell}, \tag{65}$$

where a follows from the property of two-universal hash functions and b follows from Lemma 3. Hence, $\Pr\{E_1\} \leq \varepsilon_1$ if $\ell = \Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_Y) - \log \varepsilon_1$. Hence, we have demonstrated a protocol that achieves the given error bound.

B. Proof of lower bounds in Theorem 3

Let $(e_\mathcal{X}, e_\mathcal{Y}, R) \in \Lambda^P_\varepsilon(\mathcal{X} \times \mathcal{Y} \to \mathcal{C}_\mathcal{X} \times \mathcal{C}_\mathcal{Y})$. Then there must exist a realization $R = r$ such that the error bound in (14) is met. We first prove (27).

$$\ell^\varepsilon_{\mathrm{d-enc}}(X) \stackrel{a}{=} H_0[e_\mathcal{X}(X, r)] \tag{66}$$
$$\stackrel{b}{\geq} H_0[e_\mathcal{X}(X, r)|Y] \tag{67}$$
$$\stackrel{c}{=} H_0[e_\mathcal{X}(X, r), e_\mathcal{Y}(Y, r)|Y] \tag{68}$$
$$\stackrel{d}{\geq} H_0\left\{g_\mathcal{X}\left[e_\mathcal{X}(X, r), e_\mathcal{Y}(Y, r), r\right]|Y\right\} \tag{69}$$
$$= H_0[\hat{X}(r)|Y] \tag{70}$$
$$\stackrel{e}{\geq} H^\varepsilon_0(X|Y), \tag{71}$$

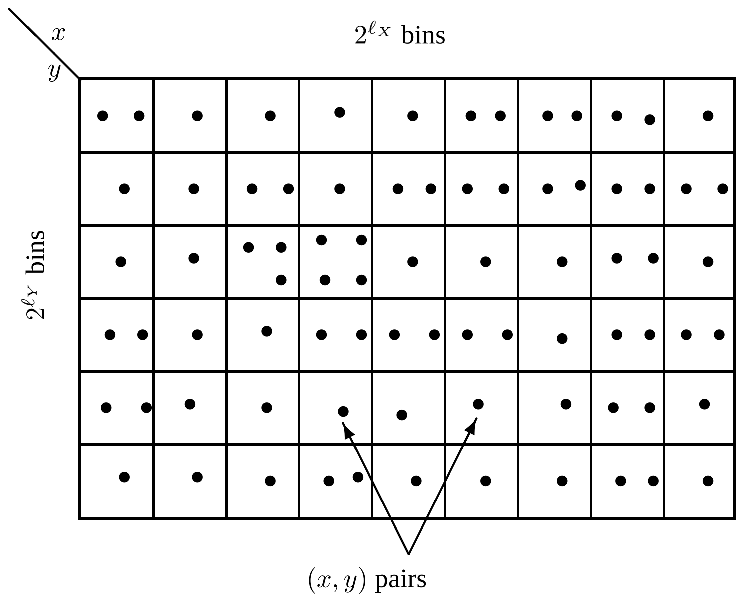

FIG. 1: The bining used in the Slepian-Wolf in Ref. [18] is used here as well.

where a follows from the definition of $\ell^{\varepsilon}_{\mathrm{d-enc}}(X)$, b follows since conditioning reduces H_0, c is an identity that is easily derived from the definition of H_0, d follows since taking a function reduces H_0, and e follows since $\Pr\{\hat{X}(r) \neq X\} \leq \varepsilon$ and from (6).

The proof for (28) is similar and is omitted. We now prove (29).

$$\ell^{\varepsilon}_{\mathrm{d-enc}}(X) + \ell^{\varepsilon}_{\mathrm{d-enc}}(Y) = H_0[e_{\mathcal{X}}(X, r)] + H_0[e_{\mathcal{Y}}(Y, r)] \tag{72}$$

$$\stackrel{a}{\geq} H_0[e_{\mathcal{X}}(X, r), e_{\mathcal{Y}}(Y, r)] \tag{73}$$

$$\geq H_0\left\{g\left[e_{\mathcal{X}}(X, r), e_{\mathcal{Y}}(Y, r), r\right]\right\} \tag{74}$$

$$= H_0[\hat{X}(r), \hat{Y}(r)] \tag{75}$$

$$\geq H^{\varepsilon}_0(X, Y), \tag{76}$$

where a follows from the sub-additivity of the Rényi entropy.

C. Proof of achievability in Theorem 3

We show the existence of a protocol for distributed encoding (again based on two-universal hash functions) and is given by the following steps.

1. Let $e_{\mathcal{X}} : \mathcal{X} \times \mathcal{R} \to \mathcal{C}_{\mathcal{X}}$, $e_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{R} \to \mathcal{C}_{\mathcal{Y}}$ be two-universal hash functions and let $\mathcal{C}_{\mathcal{X}} = \{1, 2, \cdots, 2^{\ell_{\mathcal{X}}}\}$ and $\mathcal{C}_{\mathcal{Y}} = \{1, 2, \cdots, 2^{\ell_{\mathcal{Y}}}\}$. Alice takes $x \in \mathcal{X}$, generates R (known both to Bob and Charlie), and sends $i = e_{\mathcal{X}}(x, R)$ to Charlie. Similarly, Bob takes $y \in \mathcal{Y}$ and sends $j = e_{\mathcal{Y}}(y, R)$ to Charlie. See Fig. 1 for an illustration of the encoding.

2. Charlie passes the received indices (i,j) to the decoder $g : \mathcal{C}_{\mathcal{X}} \times \mathcal{C}_{\mathcal{Y}} \times \mathcal{R} \to \mathcal{X} \times \mathcal{Y}$ that outputs $(\hat{x},\hat{y}) = (x,y)$ if there is only one pair (x,y) such that $e_{\mathcal{X}}(x,R) = i$, $e_{\mathcal{Y}}(y,R) = j$ and $(x,y) \in \mathcal{A}_\delta(P_{XY})$. In all other cases, it declares an error.

Let $\varepsilon_0, \varepsilon_1, \varepsilon_2, \varepsilon_3 \in \mathbb{R}^+$ and $\sum_{k=0}^{3} \varepsilon_k \leq \varepsilon$. The probability of error of the above protocol is computed by defining the following events.

$$
\begin{aligned}
E_0 &:= \{(X,Y) \notin \mathcal{A}_\delta(P_{XY})\}, && (77)\\
E_1 &:= \{\exists\, \acute{x} \neq X : e_{\mathcal{X}}(\acute{x},R) = e_{\mathcal{X}}(X,R), (\acute{x},Y) \in \mathcal{A}_\delta(P_{XY})\}. && (78)\\
E_2 &:= \{\exists\, \acute{y} \neq Y : e_{\mathcal{Y}}(\acute{y},R) = e_{\mathcal{Y}}(Y,R), (X,\acute{y}) \in \mathcal{A}_\delta(P_{XY})\}. && (79)\\
E_{12} &:= \{\exists(\acute{x},\acute{y}) : (\acute{x},\acute{y}) \neq (X,Y), e_{\mathcal{X}}(\acute{x},R) = e_{\mathcal{X}}(X,R), e_{\mathcal{Y}}(\acute{y},R) = e_{\mathcal{Y}}(Y,R), \\
&\qquad (\acute{x},\acute{y}) \in \mathcal{A}_\delta(P_{XY})\}. && (80)
\end{aligned}
$$

The probability of error is given by the probability of the union of these events and we upper bound that by the union bound as

$$
\begin{aligned}
P_e &= \Pr\{E_0 \cup E_1 \cup E_2 \cup E_{12}\} && (81)\\
&\leq \Pr\{E_0\} + \Pr\{E_1\} + \Pr\{E_2\} + \Pr\{E_{12}\}. && (82)
\end{aligned}
$$

For any $\varepsilon_0 \geq 0$, we could choose a $\delta \geq 0$ such that $\Pr\{E_0\} \leq \varepsilon_0$. We now bound $\Pr\{E_1\}$ as follows.

$$
\begin{aligned}
\Pr\{E_1\} &= \Pr\{\exists\, \acute{x} \neq X : e_{\mathcal{X}}(\acute{x},R) = e_{\mathcal{X}}(X,R), (\acute{x},Y) \in \mathcal{A}_\delta(P_{XY})\} && (83)\\
&= \sum_{x,y} P_{XY}(x,y) \Pr\{\exists\, \acute{x} \neq x : e_{\mathcal{X}}(\acute{x},R) = e_{\mathcal{X}}(x,R), (\acute{x},y) \in \mathcal{A}_\delta(P_{XY})\} && (84)\\
&= \sum_{x,y} P_{XY}(x,y) \sum_{\acute{x}\neq x, (\acute{x},y)\in\mathcal{A}_\delta(P_{XY})} \Pr\{e_{\mathcal{X}}(\acute{x},R) = e_{\mathcal{X}}(x,R)\} && (85)\\
&\overset{a}{\leq} \sum_{x,y} P_{XY}(x,y) 2^{-\ell_{\mathcal{X}}} |\mathcal{A}_\delta(P_{X|Y=y})| && (86)\\
&\overset{b}{\leq} 2^{\Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_Y) - \ell_{\mathcal{X}}}, && (87)
\end{aligned}
$$

where a follows from the property of two-universal hash function and b follows from Lemma 3. Hence, $\Pr\{E_1\} \leq \varepsilon_1$, if $\ell_{\mathcal{X}} \geq \Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_Y) - \log\varepsilon_1$. Similarly, we could show that $\Pr\{E_2\} \leq \varepsilon_2$, if $\ell_{\mathcal{Y}} \geq \Xi^\delta_{\max}(P_{XY}) - \Xi^\delta_{\min}(P_X) - \log\varepsilon_2$ and $\Pr\{E_{12}\} \leq \varepsilon_3$, if $\ell_{\mathcal{X}} + \ell_{\mathcal{Y}} \geq \Xi^\delta_{\max}(P_{XY}) - \log\varepsilon_3$.

D. Proofs of Lemmas 1 and 2

The proofs of Lemmas 1 and 2 follow straightforwardly if we could invoke Lemma 5, which we can if we can show that δ is not zero for large enough n. It is here that we shall need the following remarkable result by Holenstein and Renner [16] giving explicit bounds.

Theorem 4. *(Holenstein and Renner [16]) Let $P_{\mathbf{X}_1^n\mathbf{Y}_1^n} := P_{X_1Y_1}\cdots P_{X_nY_n}$ be a probability distribution over $\mathcal{X}^n \times \mathcal{Y}^n$. Then, for any $\delta \in [0, \log|\mathcal{X}|]$ and $\mathbf{x},\mathbf{y}$ chosen according to $P_{\mathbf{X}_1^n\mathbf{Y}_1^n}$*

$$
\Pr_{\mathbf{x},\mathbf{y}}\left[-\log(P_{\mathbf{X}_1^n|\mathbf{Y}_1^n}(\mathbf{x},\mathbf{y})) \geq H(\mathbf{X}_1^n|\mathbf{Y}_1^n) + n\delta\right] \leq \varepsilon \qquad (88)
$$

and, similarly

$$\Pr_{\mathbf{x},\mathbf{y}}\left[-\log(P_{\mathbf{X}_1^n|\mathbf{Y}_1^n}(\mathbf{x},\mathbf{y})) \leq H(\mathbf{X}_1^n|\mathbf{Y}_1^n) - n\delta\right] \leq \varepsilon \tag{89}$$

where $\varepsilon = 2^{-\frac{n\delta^2}{2\log^2(|\mathcal{X}|+3)}}$.

Let us now assume that we want to ensure that for $\delta = \varepsilon_0$,

$$\Pr\{\mathbf{X}_1^n \notin A_\delta(P_{\mathbf{X}_1^n})\} \leq \varepsilon_0. \tag{90}$$

Recall that $X_1, X_2, ...$ are i.i.d. and distributed according to P_X. Then it is sufficient to ensure that

$$\Pr_{\mathbf{x}}\left[-\log(P_{\mathbf{X}_1^n}(\mathbf{x})) \geq nH(X_1) + n\delta\log|\mathcal{X}|\right] \leq \frac{\varepsilon_0}{2}, \tag{91}$$

$$\Pr_{\mathbf{x}}\left[-\log(P_{\mathbf{X}_1^n}(\mathbf{x})) \leq nH(X_1) - n\delta\log|\mathcal{X}|\right] \leq \frac{\varepsilon_0}{2}, \tag{92}$$

where we have taken out the term inside $|\cdot|$ to get upper bounds to the probabilities. Now invoking the result by Holenstein and Renner [16], if we choose $\delta = \varepsilon_0$, then for all $n \geq n_0(\varepsilon_0)$, where

$$n_0(\varepsilon_0) := \sqrt{-\frac{2\log(\varepsilon_0/2)\ \log^2(|\mathcal{X}|+3)}{\log^2(|\mathcal{X}|)\,\varepsilon_0^2}}, \tag{93}$$

the error bound in (90) is met. Note that we could choose $\delta = 0$ for $n < n_0(\varepsilon_0)$. Hence, both the sequences $\{H_{-\infty}^{\delta}(X_1^n)/n\}$ and $\{H_{-\infty}^{\varepsilon_0}(X_1^n)/n\}$ are equal $\forall\ n \geq n_0(\varepsilon_0)$. Hence,

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{H_{-\infty}^{\delta}(X_1^n)}{n} = \lim_{\varepsilon\to 0}\lim_{n\to\infty}\frac{H_{-\infty}^{\varepsilon_0}(X_1^n)}{n} = \lim_{\varepsilon_0\to 0}\lim_{n\to\infty}\frac{H_{-\infty}^{\varepsilon_0}(X_1^n)}{n}. \tag{94}$$

Now Lemma 5 can be invoked to yield the desired answer. It is straightforward to extend the above argument for the i.i.d. sequence of a pair of random variables (X, Y).

VI. CONCLUSIONS AND ACKNOWLEDGEMENT

In conclusion, we have given the one-shot bounds for the Slepian-Wolf protocol for distributed encoding of correlated sources [18]. We constructed a protocol for distributed encoding using two-universal hash functions and this gave the achievability region for the rate-pairs. We show that the bounds are asymptotically tight and yield the same rate region as the Slepian-Wolf. Our protocol is based on defining a set based on Smooth Rényi entropies reminiscent of the typical set in the asymptotic case. The authors are not aware of any pre-existing one-shot results for the multi-party scenario for the classical case.

The authors wish to thank R. Renner for giving his helpful comments on the subject.

---

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Sys. Tech. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2nd ed., 2006.

[3] R. Renner and S. Wolf, "Smooth Rényi entropy and applications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Chicago, IL, USA), June 2004.

[4] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in Cryptology—ASIACRYPT 2005, Lecture Notes in Computer Science*, pp. 199–216, Springer-Verlag, 2005.

[5] C. Cachin and U. Maurer, "Smoothing probability distributions and smooth entropy (extended abstract)," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Ulm, Germany), June 1997.

[6] R. Renner, S. Wolf, and J. Wullschleger, "The single-serving channel capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Seattle, WA, USA), July 2006.

[7] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Seoul, Korea), June 2009.

[8] R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," in *Proc. Theory Crypt. Conf. (TCC)*, vol. 3378, pp. 407–425, Lecture Notes in Computer Science, Springer-Verlag, 2005.

[9] F. Dupuis, P. Hayden, and K. Li, "A father protocol for quantum broadcast channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 2946–2956, June 2010.

[10] J. M. Renes and R. Renner, "Noisy channel coding via privacy amplification and information reconciliation," *IEEE Trans. Inf. Theory*, vol. 57, pp. 7377–7385, Nov. 2011.

[11] M. Berta, M. Christandl, and R. Renner, "The quantum reverse Shannon theorem based on one-shot information theory," *Commun. Math. Phys.*, vol. 306, pp. 579–615, Sept. 2011.

[12] N. Datta and M.-H. Hsieh, "The apex of the family tree of protocols: optimal rates and resource inequalities," *New J. Phys.*, vol. 13, p. 093042, Sept. 2011.

[13] N. Datta and R. Renner, "Smooth Rényi entropies and the quantum information spectrum," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2807–2815, 2009.

[14] R. König, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Trans. Inf. Theory*, vol. 55, pp. 4337–4347, Sept. 2009.

[15] B. Schoenmakers, J. Tjoelker, P. Tuyls, and E. Verbitskiy, "Smooth Rényi entropy of ergodic quantum information sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, (Nice, France), June 2007.

[16] T. Holenstein and R. Renner, "On the randomness of independent experiments," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1865–1871, April 2011.

[17] J. M. Renes and R. Renner, "One-shot classical data compression with quantum side information and the distillation of common randomness or secret keys." arXiv:1008:0452.

[18] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.

[19] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math Stat. Prob.*, pp. 547–561, 1960.

[20] C. Bennett, G. Brassard, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inf. Theory*, vol. 41, pp. 1915–1923, 1995.